SAM Distributed Analysis Cluster Caching

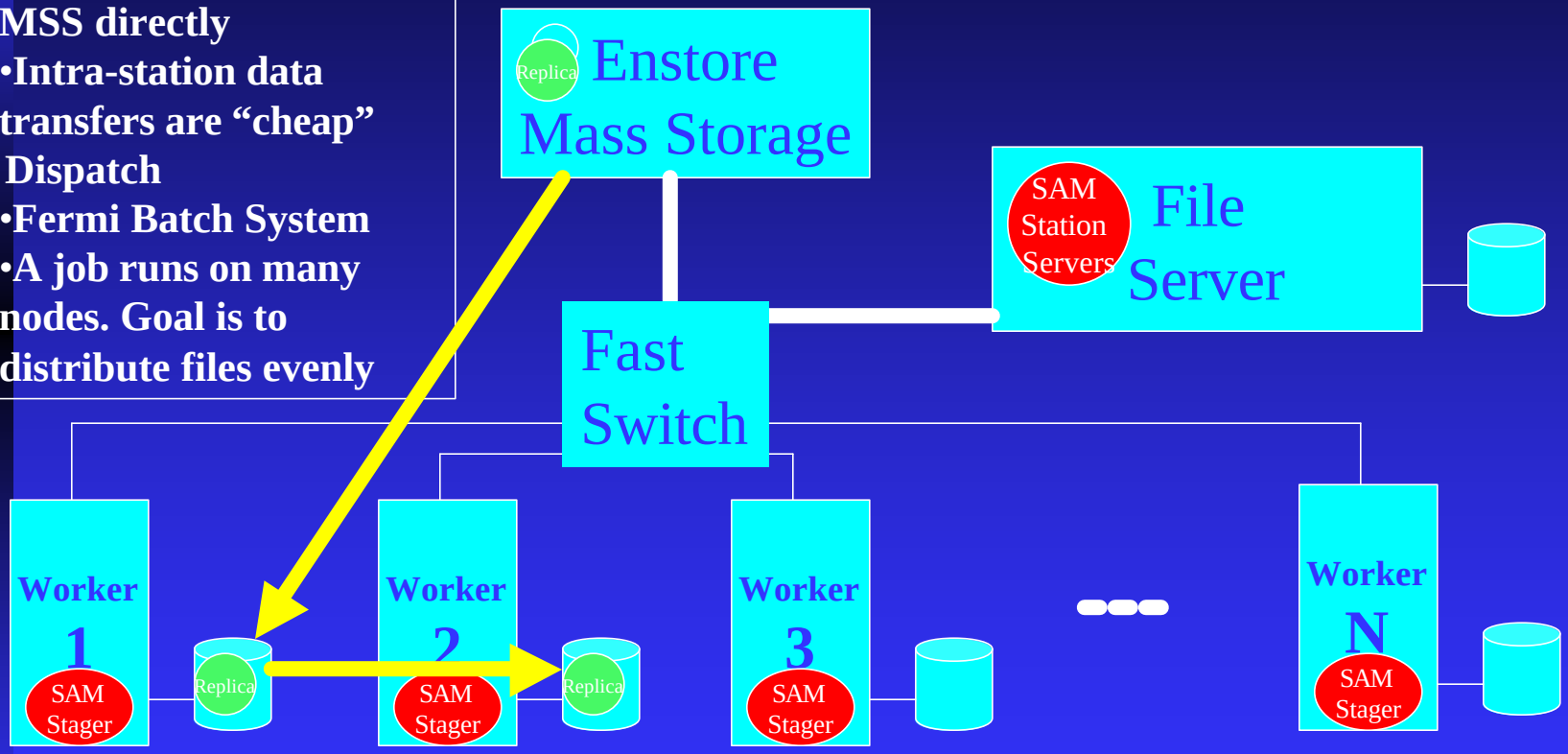
Lee Lueking

CD Strategy Meeting

February 6, 2002

SAM Station: Distributed Reconstruction Farm

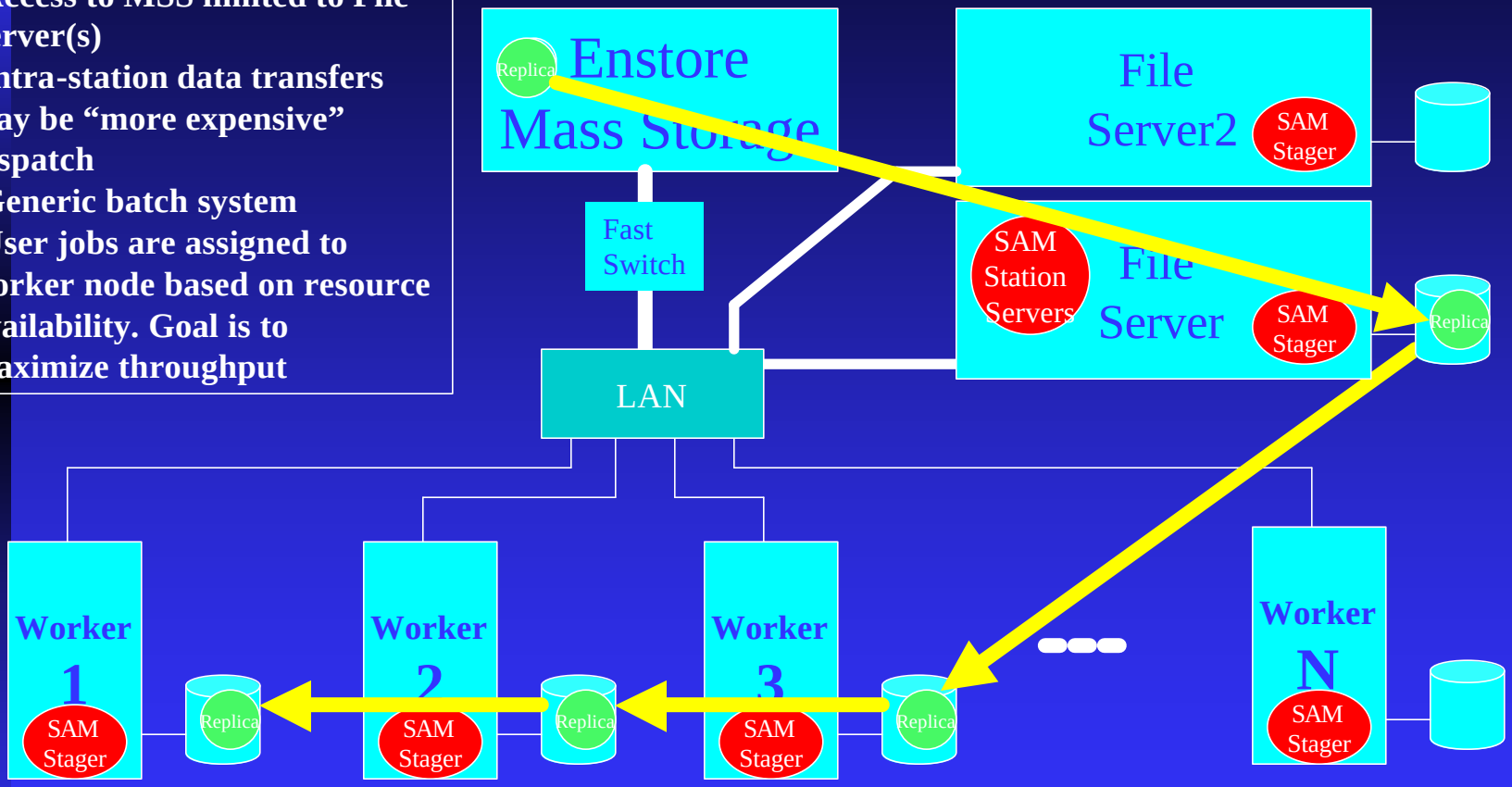
- **Network**
  - **Each worker accesses MSS directly**
  - **Intra-station data transfers are "cheap"**
- **Job Dispatch**
  - **Fermi Batch System**
  - **A job runs on many nodes. Goal is to distribute files evenly**



No disks are cross mounted. Worker nodes get files directly from MSS via encp. Data is moved by SAM using rcp from where it is cached to where it is needed.

SAM Station: Distributed Analysis Cluster

- **Network**
  - **Access to MSS limited to File Server(s)**
  - **Intra-station data transfers may be "more expensive"**
- **Job Dispatch**
  - **Generic batch system**
  - **User jobs are assigned to worker node based on resource availability. Goal is to maximize throughput**



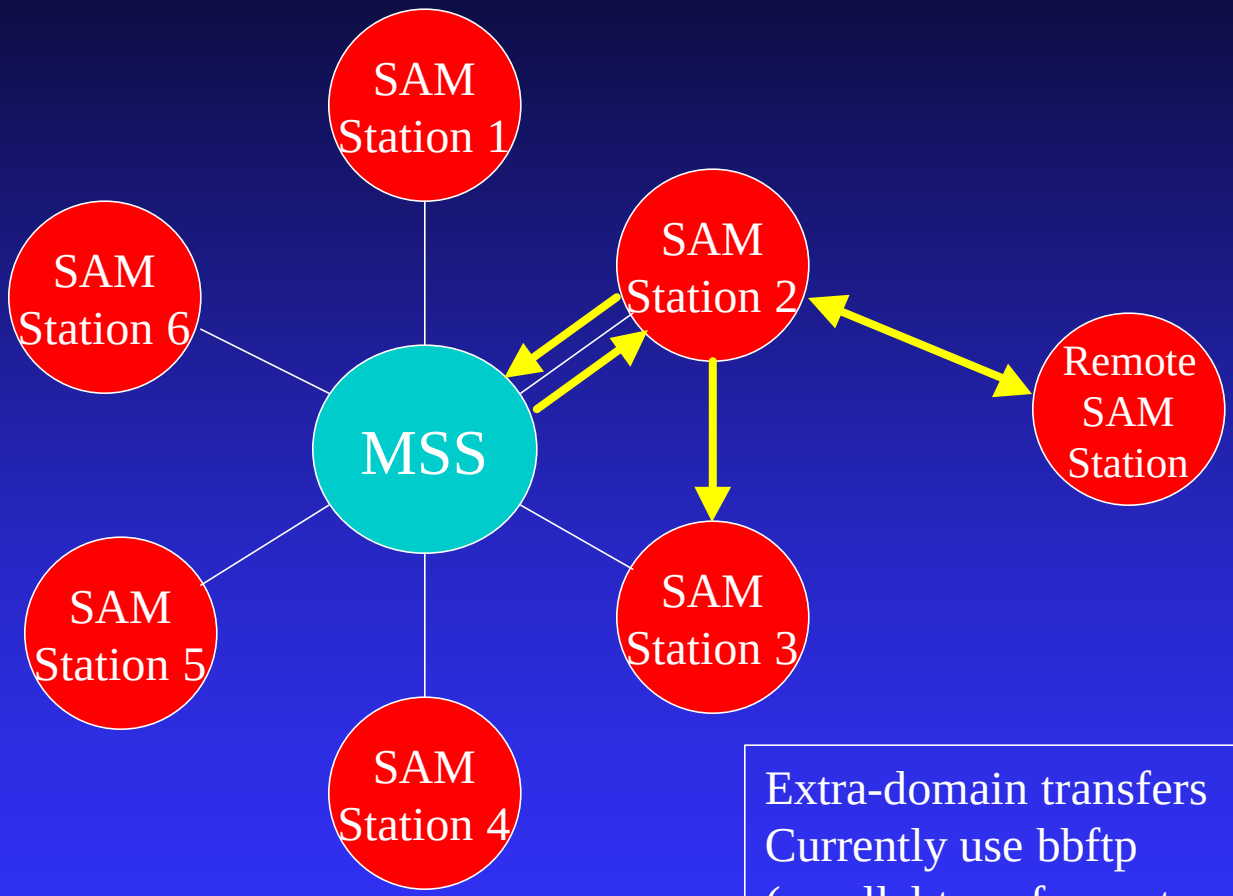
No disks are cross mounted. File Server node(s) get data directly from MSS via encp. Data is replicateded by SAM using rcp from where it is cached to where it is needed.

Station Considerations

- **Resource Co-allocation of data and processing resources:**
  - **Temporal locality – Data and Processing are available at the same time (implemented).**
  - **Spatial locality – Data and Processing are available at the same place (not implemented yet).**
  - **SAM implements through "batch adapter", dependent on native batch system.**
- **Dedicated file server(s):**
  - **Highly desirable. Needed to control access to MSS.**
  - **Load balancing among multiple file servers does not exist. Assumption is that "random" will work to get started.**
  - **More complexity will be added as it is understood.**
- **Network:**
  - **Currently, station implementation assumes intra-station data transfers are cheap.**
  - **Ultimately, load balancing of all station resources is desired including network, CPU, cache, etc.**

Site and Global Issues

Station Configuration
- Replica location
  - Prefer
  - Avoid
- Forwarding
  - File stores can be forwarded through other stations
- "Routing" (now)
  - Parasitic Stagers
- Routing (soon)
  - Can set up routes for replica transfers



Extra-domain transfers
Currently use bbftp (parallel transfer protocol)

Questions

- Access Patterns:What are the anticipated access patterns for each facility?
  - What is the CPU/IO ratio?
    - High – Reconstruction
    - Medium
    - Low – Thumbnail mining
  - What is the cache hit ratio (data requested already in the cache)/(total data requested)?
    - High – Same files being reprocessed repeatedly
    - Medium
    - Low – New data constantly
  - How should Jobs be dispatched to most effectively use the resources (achieve the highest throughput)?

Questions

- **Scalability: How scalable is each solution?**
- **Reliability: Are there single points of failure?**
- **Security: What are the security issues and how are they addressed?**
- **Usability:**
  - **Does the solution provide maximum usability to the users**
  - **Does it make the use of the D0 data handling model (SAM) to its fullest?**
- **Network:**
  - **How efficient is the intra-station network?**
  - **What network resources are needed for extra-station transfers?**
  - **How does station activity impact outside users?**
- **Costs: What are the real costs for each deployment?**
  - **Hardware including networking**
  - **Administrative and monitoring**